



MARS 2024

PUCES ÉLECTRONIQUES ET IA

Le marché des puces premium s'envole, profitant aux acteurs en place dans un contexte géopolitique mouvant

PUCES ÉLECTRONIQUES ET IA

Le marché des puces électroniques premium s'envole, profitant aux acteurs en place dans un contexte géopolitique mouvant

Marché cyclique ayant récemment traversé une période difficile, le domaine des semi-conducteurs connaît un nouvel essor, en particulier sur son segment haut de gamme. Dynamisées par les progrès de l'intelligence artificielle, notamment dans les modèles de langage (IA générative), les ventes de puces électroniques premium accompagnent la montée des besoins en capacités de calcul et en traitement de données.

Visible sur l'ensemble de la filière, cette émulation ne bénéficie toutefois qu'à un petit nombre d'acteurs, en raison de la forte concentration du secteur sur les segments les plus haut de gamme. Des fonderies aux *fabs* en passant par les fournisseurs d'équipements, chaque maillon de la chaîne de valeur se trouve dans une situation quasi-monopolistique.

Face à cet essor, la concurrence tend cependant à se renforcer, que ce soit entre industriels ou entre États. La dualité des usages – civils et militaires – des puces les plus sophistiquées conduit à une guerre commerciale sino-américaine, dont les effets se révèlent contrastés pour les acteurs impactés. Les questions environnementales constituent une autre problématique pour un secteur fortement consommateur de ressources.



DANS CE DOSSIER

RECOMMANDATIONS	4
-----------------------	---

LA MONTÉE EN PUISSANCE DE L'IA DYNAMISE LE MARCHÉ DES SEMI-CONDUCTEURS	5
---	---

L'intelligence artificielle se déploie dans une multitude de domaines5

L'écosystème des IA génératives pousse à la hausse la demande
en semi-conducteurs.....7

Après quelques turbulences, le marché retrouve une croissance soutenue8

La consommation des équipements, un défi économique et environnemental9

LA FILIÈRE DES PUCES ÉLECTRONIQUES PROFITE DE CE NOUVEL ÉLAN	10
--	----

Une chaîne de valeur structurée autour de quatre catégories d'acteurs principales .. 10

Nvidia, grand gagnant de l'essor de l'IA..... 11

Des alternatives émergent pour réduire l'hégémonie de Nvidia..... 13

Parmi les fonderies, TSMC apparaît comme le leader incontesté
des puces premium 16

Les progrès technologiques se poursuivent dans les équipements de lithographie.... 17

Entre opportunités et menaces, le contexte géopolitique guide les mutations
du secteur..... 18

PRINCIPALES SOURCES UTILISÉES.....	23
------------------------------------	----

RECOMMANDATIONS

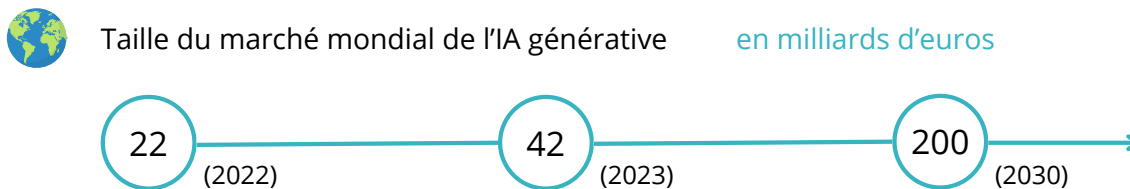
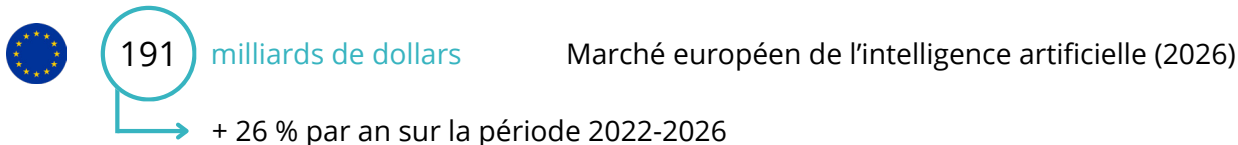
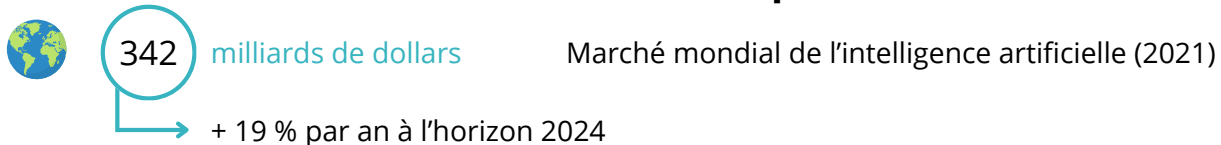
Ce qu'il faut retenir

- **L'accélération des technologies basées sur l'intelligence artificielle** et la diffusion de solutions afférentes dans le monde professionnel constituent **de puissants moteurs de développement des puces électroniques** de haute performance.
- Si le potentiel semble considérable, le secteur pris dans sa globalité n'est **pas à l'abri d'un retournement de tendance** sur les puces moins sophistiquées. Les acteurs les plus exposés à ces segments pourraient subir **des déconvenues commerciales et financières**. Certains experts regardent même avec scepticisme l'envolée du cours boursier des leaders du secteur.
- **La multiplication des projets d'usines**, l'alourdissement de la facture énergétique des data centers ou encore le ralentissement économique pourraient conduire à **une situation de surproduction** de puces électroniques.
- De l'amont à l'aval, **chaque maillon de la chaîne de valeur s'avère fortement concentré**, en particulier sur le segment haut de gamme. Un acteur dominant se place en général dans **une situation proche du monopole**, combinant la maîtrise des technologies les plus avancées et des capacités de production supérieures.
- Conscients de leur dépendance à l'égard de Nvidia, principal fournisseur actuel, **les GAFAM et autres hyperscalers** cherchent à diversifier leurs approvisionnements, voire à **concevoir leurs propres puces** et à s'adresser directement aux fonderies pour **limiter les dépenses considérables** liées au déploiement de leurs modèles d'IA.
- **De nouveaux acteurs émergent sur le segment des fabless** afin de proposer de nouvelles puces à ces grands groupes. Modèle flexible, nécessitant relativement peu de capitaux, **l'approche "sans usine" s'avère propice pour se positionner** comme nouvel entrant. L'inférence, soit le traitement de données lié à l'utilisation des IA, se montre également **plus accessible que l'apprentissage algorithmique** en termes de capacités de calcul.
- Concernant les politiques gouvernementales, plusieurs approches se dessinent. Les États européens, consommateurs de puces pour l'industrie ou l'automobile, souhaitent **réduire leurs importations en y substituant une production locale**. Les pays fabricants comme Taïwan ou la Corée du Sud viennent soutenir le secteur, exportateur et stratégique, pour **conserver ou accroître leurs parts de marché**. Leaders dans la commercialisation, les États-Unis combinent ces différentes stratégies avec des mesures de guerre commerciale (embargos) s'inscrivant dans **une logique de rivalité technologique** avec la Chine.
- Ce retour des questions de souveraineté entraîne diverses répercussions. En tant que **fabless, miser sur une production locale** peut permettre de se démarquer de ses concurrents. Pour les fondeurs, **s'implanter au plus près des marchés** prend de l'importance auprès des clients et des gouvernements, plus attentifs à la résilience de leur approvisionnement.

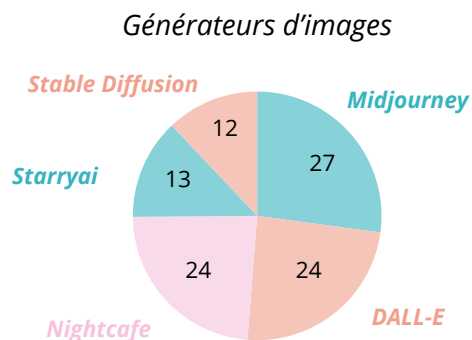
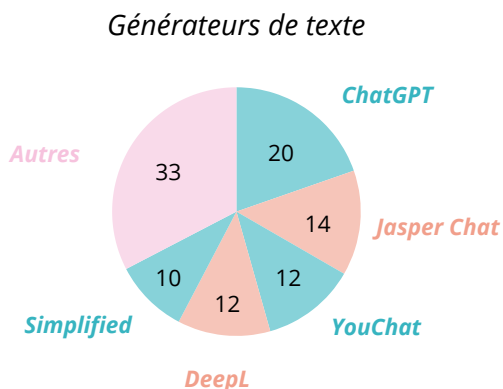
LA MONTÉE EN PUISSANCE DE L'IA DYNAMISE LE MARCHÉ DES SEMI-CONDUCTEURS

L'intelligence artificielle se déploie dans une multitude de domaines

L'IA, un marché en expansion



Parts de marché des principaux outils d'IA génératives en %



Traitement IndexPresse. Sources : IDC, Statista

LA MONTÉE EN PUISSANCE DE L'IA DYNAMISE LE MARCHÉ DES SEMI-CONDUCTEURS

“Nous sommes entrés dans l'ère de la décision et du travail augmentés par l'IA dans tous les domaines fonctionnels de l'entreprise. La création et l'utilisation responsables de solutions d'IA capables de détecter, prédire, répondre et s'adapter rapidement deviennent des impératifs commerciaux.”

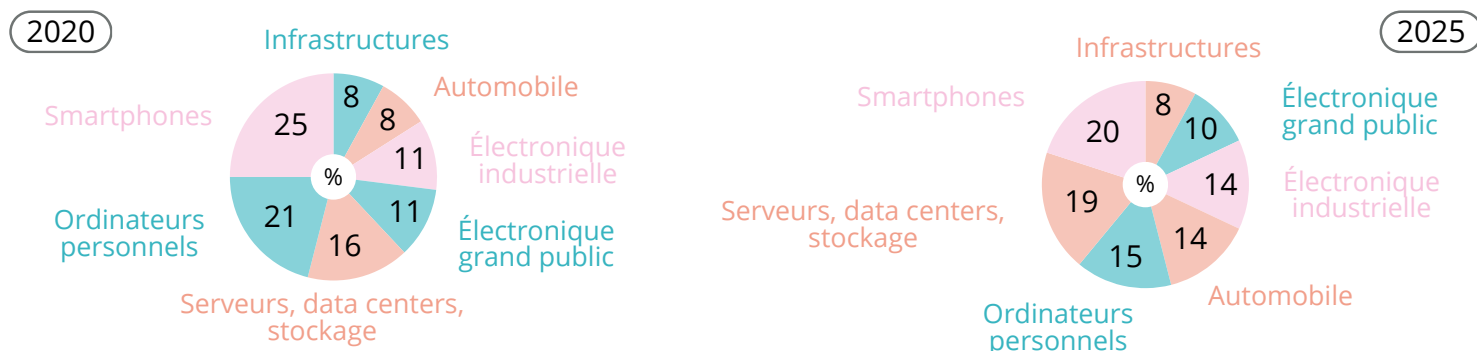
Ritu Jyoti, vice-présidente de recherche sur l'IA et l'automatisation chez IDC, 2021

- Des outils recourant à l'intelligence artificielle se mettent en place **dans de nombreux secteurs, notamment sous la forme de logiciels**. Il devrait s'agir du segment le plus dynamique du marché en Europe, avec une croissance de 32 % par an entre 2022 et 2026 (source : IDC).
- Cette diffusion de l'IA permet d'**accroître la productivité pour des tâches variées** : recherche d'informations, gestion administrative, traitement de données, modélisation industrielle, relation client...
- L'**essor de l'IA générative propulse le domaine dans une nouvelle dimension**, avec une hausse considérable du nombre d'utilisateurs, notamment au sein du grand public. Il atteignait 250 millions de personnes à l'échelle mondiale en 2023, et pourrait s'élever à 700 millions à l'horizon 2030 (source : Statista).
- Création de textes ou d'images, incorporation dans les robots conversationnel virtuels et, à l'avenir, physiques : **les grands modèles de langage (LLM) s'intègrent dans une profusion d'usages** et étendent encore le potentiel de l'IA.

L'écosystème des IA génératives pousse à la hausse la demande en semi-conducteurs

Le segment des serveurs prend une place croissante sur le marché

Parts de marché des usages finaux des puces électroniques, en valeur



Traitement IndexPresse. Source : Statista

- Les grands acteurs du numérique (GAFAM) se lancent dans le domaine de l'IA générative, parfois **en s'appuyant sur des start-up**. Cette alliance s'explique par **les colossales ressources nécessaires** pour faire fonctionner les modèles d'IA génératives, que seuls de grands groupes peuvent supporter. Le lancement de ChatGPT a ainsi coûté **700 000 dollars par jour** (source : *Challenges*).

- **Microsoft a financé OpenAI** à hauteur de 1 milliard de dollars en 2019, avant de racheter 49 % des parts de la société en 2023 pour 10 milliards de dollars.

- **Google** a investi 300 millions de dollars en 2023 pour acquérir **10 % du capital de la start-up Anthropic**. Le groupe a rajouté 500 millions de dollars quelques mois plus tard, et s'est engagé à apporter 1,5 milliard dans le futur. Google a également noué **un partenariat avec la société française Mistral AI**.

- **Amazon a lui aussi financé Anthropic** pour un montant de 1,25 milliard de dollars, et pourrait soutenir la start-up jusqu'à 4 milliards de dollars.

- **Des regroupements de laboratoires** sont par ailleurs effectués au sein des groupes pour renforcer la recherche dans le domaine. C'est le cas

chez Google où les centres DeepMind et Google Brain ont fusionné, et chez Meta avec le rapprochement de sa division de recherche fondamentale FAIR avec son laboratoire GenAI.

- **Les outils se multiplient** et entrent en concurrence. Outre son assistant Copilot, Microsoft s'appuie sur ChatGPT d'OpenAI pour la génération de texte. Google dispose de son robot conversationnel Bard et de son modèle Gemini. Meta a lancé Llama sur le même créneau, tandis qu'Amazon propose Q à la clientèle professionnelle.

- **Le type de contenu proposé se diversifie** également. OpenAI a ainsi développé DALL-E pour la création d'images, et Sora dans celle de vidéos. De son côté, Meta a mis au point la plateforme Imagine (basée sur son modèle Emu) pour la génération d'images. Gemini de Google se veut polyvalent, capable de créer différentes sortes de contenus.

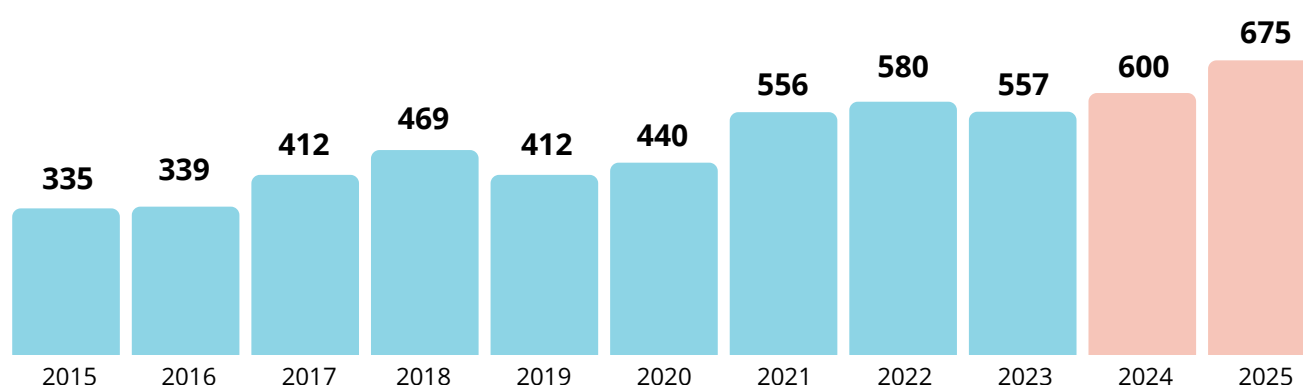
- Ce dynamisme de l'IA générative conduit à **une forte demande pour les puces électroniques** utilisées dans les data centers et autres équipements fournissant la puissance de calcul des modèles génératifs. À l'avenir, les utilisations industrielles vont ainsi prendre le dessus sur les usages grand public.



Après quelques turbulences, le marché retrouve une croissance soutenue

Le marché des semi-conducteurs efface l'effet Covid et affiche des perspectives prometteuses

Ventes mondiales de semi-conducteurs, en milliards de dollars



Traitement IndexPresse. Sources : Statista, *Investir*

- Les dysfonctionnements des chaînes logistiques liées à la crise de Covid-19 ont profondément affecté la santé du secteur à partir de 2020. Les fabricants ont annoncé des augmentations de prix successives (+20 % en 2021 pour TSMC, +15 % à +20 % chez Samsung à l'été 2022). **L'inflation** a pénalisé les ventes de matériel informatique, puis **la hausse des frais énergétiques** a diminué la rentabilité du secteur. La pénurie aurait causé 500 milliards de dollars de pertes aux entreprises de semi-conducteurs et à leurs clients (source : *Investir*). L'indice boursier des trente plus grands acteurs américains du domaine, le Philadelphia Semiconductor Sector, est passé d'environ 4 000 points fin 2021 à 2 600 points fin 2022.

- Le potentiel de marché apporté par l'essor des IA génératives a toutefois permis de redynamiser le secteur. **Déjà accrus par les changements de pratiques numériques** des entreprises, comme le recours au cloud ou la généralisation du télétravail et donc des outils de visioconférence, les

besoins en équipements informatiques augmentent encore avec l'implantation de l'IA dans les habitudes de travail.

- Le marché profite en outre du fait que les puces électroniques concernées par l'IA, généralement les plus petites (entre 3 et 10 nanomètres) **sont les plus performantes et donc les plus valorisées**. Les processeurs les plus puissants du concepteur américain Nvidia, les H100, sont ainsi **vendus à plus de 40 000 dollars l'unité**. "Plus difficiles à trouver que de la drogue", ironisait le milliardaire Elon Musk, elles bénéficient d'une demande très forte de la part des acteurs de l'IA.

Le fondeur TSMC a quant à lui réalisé un tiers de son chiffre d'affaires dans l'informatique en 2020, contre seulement 3 % dans l'automobile, une industrie consommatrice de puces moins sophistiquées. Il génère ainsi **deux à trois fois plus de chiffre d'affaires** que ses concurrents par *wafer* (galette de silicium à la base de la fabrication des semi-conducteurs).

La consommation des équipements, un défi économique et environnemental

“Il y a trois principaux postes de consommation [d'eau]. Le premier est le rinçage des plaquettes de puces après chaque opération. (...) Viennent ensuite le refroidissement de certains équipements de production [et la station de traitement des effluents].”

Guy Dubois, ancien manager chez STMicroelectronics

“Les allers-retours entre un processeur et sa mémoire consomment 100 à 1 000 fois plus d'énergie qu'une simple opération d'addition.”

Manuel Le Gallo, chercheur chez IBM Research Group, septembre 2023

- L'impact environnemental de la fabrication de semi-conducteurs et de l'utilisation de ces derniers, en particulier dans le cadre de l'IA, peut constituer une problématique majeure. Il s'agit notamment de **limiter l'envolée de la facture énergétique** liée à une consommation considérable d'électricité, tout en **réduisant les besoins en eau** des usines de fabrication. Dans le cas contraire, des contestations sur l'usage de la ressource hydrique pourraient pénaliser le secteur.

- Chez le fabricant franco-italien STMicroelectronics, **des efforts ont été réalisés pour diminuer la consommation d'eau**. Les équipements les plus récents affichent ainsi des besoins vingt fois inférieurs à ceux datant des années 2000, selon l'entreprise. La consommation a été abaissée de 41 % par puce produite entre 2016 et 2022. Les objectifs de recyclage de l'eau visent à atteindre un taux de 60 %, contre 43 % en 2022 et 49 % en 2023.

Ces efforts en matière de responsabilité environnementale se montrent **nécessaires face aux prévisions d'accroissement de la consommation d'eau**. Cette dernière pourrait atteindre entre 800 et 1 200 mètres cubes d'eau par heure dans son usine de Crolles (Isère) à partir de 2026, contre 600 en 2023. Avec 4,2 millions de mètres cubes d'eau consommés en 2022, l'usine subit des critiques de la part d'habitants locaux.

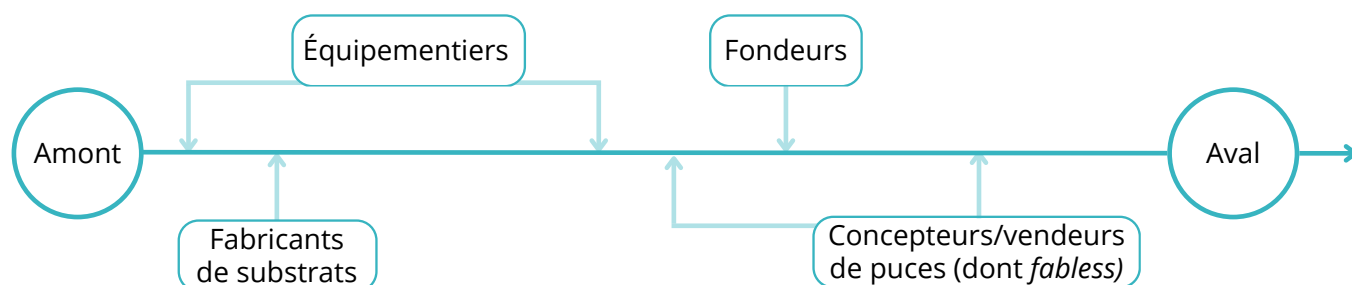
- L'IA représente également **une source de consommation électrique majeure**. Le seul entraînement de ChatGPT-3 aurait entraîné une consommation de 190 MWh d'électricité en 2020, soit l'équivalent de la consommation annuelle d'environ 90 Français (source : *Industrie & Technologies*). Entre les dépenses en matériel et l'alimentation électrique, les modèles d'IA génératives et les autres services cloud s'avèrent donc onéreux pour leurs fournisseurs.

- Des recherches sont menées pour réduire la consommation électrique. **La miniaturisation des semi-conducteurs** constitue une première voie. Les puces de 3 nanomètres développées par TSMC ont ainsi affiché une baisse de leur consommation énergétique de 30 % par rapport à celles de 5 nanomètres, tout en améliorant leurs performances d'environ 15 %.

- D'autres approches sont explorées, telles que **le calcul neuromorphique**. S'inspirant de la structure du cerveau, de nouvelles puces sont développées et visent à **combinaison calcul et mémoire**, des fonctions souvent séparées dans les équipements informatiques. Des gains de performances considérables pourraient être atteints en empruntant cette voie. Les puces Hermès d'IBM ou le projet européen Human Brain Project se positionnent sur le sujet.

LA FILIÈRE DES PUCES ÉLECTRONIQUES PROFITE DE CE NOUVEL ÉLAN

Une chaîne de valeur structurée autour de quatre catégories d'acteurs principales



Traitement IndexPresse.

- L'accélération de la croissance du secteur bénéficie aux acteurs déjà en place, d'autant plus que cette industrie se révèle **peu ouverte aux nouveaux entrants** : complexité technologique nécessitant des années de R&D et forte intensité capitalistique viennent conforter la place des acteurs historiques.

- La filière se répartit en quatre segments principaux :

- en amont, **les fabricants de substrats** produisent, en général à partir de silicium, le premier composant nécessaire à la création de puces électroniques. Ils fournissent cette "matière première" aux fonderies. En France, l'entreprise Soitec est positionnée sur ce segment ;

- au cœur du secteur figurent **les fondeurs, c'est-à-dire les fabricants de puces**, tels que TSMC, Intel ou STMicroelectronics. Ils peu-

vent **opérer en tant que sous-traitant**, en fabriquant des puces conçues et commercialisées par d'autres, ou **suivre une approche verticalement intégrée** ;

- **les équipementiers** fournissent les machines nécessaires à la fabrication des puces et des substrats. Le néerlandais ASML constitue le leader mondial dans le domaine ;

- **les concepteurs-vendeurs** de puces, qui s'occupent donc à la fois de l'amont (design des puces) et de l'aval (commercialisation). On trouve dans cette catégorie **les fabless**, des acteurs sans usine sous-traitant l'intégralité de la fabrication de leurs puces. C'est notamment le cas de Nvidia. **Il s'agit du modèle présentant le moins de barrières à l'entrée**, du fait de l'absence de sites de fabrication. D'autres acteurs, comme Intel ou Samsung, disposent eux de leurs propres sites de fabrication.

Nvidia, grand gagnant de l'essor de l'IA

Une progression fulgurante pour Nvidia en 2023

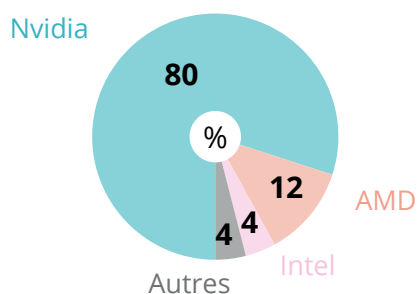
Cours de Bourse

2 200	milliards de dollars	mars 2024	} x7 en un an
1 000	milliards de dollars	mai 2023	
300	milliards de dollars	janvier 2023	
100	milliards de dollars	2019	

seules 4 autres entreprises ont atteint les 2 000 milliards de dollars de valorisation

Microsoft Google
Apple Saudi Aramco

Parts de marché dans les GPU



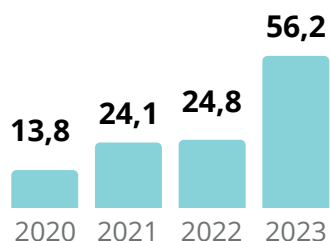
Classement des principaux acteurs de la vente de semi-conducteurs, en 2023

1	Intel	9,1 %	} parts de marché
2	Samsung	7,5 %	
3	Qualcomm	5,4 %	
4	Broadcom	4,8 %	
5	Nvidia	4,5 %	

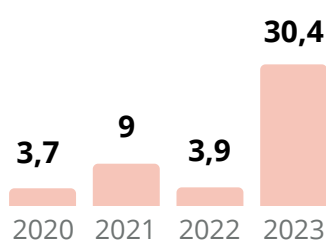
→ première fois que Nvidia entre dans le top 5

Évolution des performances économiques en milliards de dollars

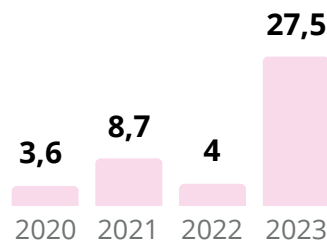
Chiffre d'affaires



Résultat opérationnel



Résultat net



Traitement IndexPresse. Sources : Challenges, Les Échos, Boursorama, Stocklytics

“Combien de temps [les grandes entreprises du numérique] vont-elles pouvoir dépenser des sommes astronomiques auprès de Nvidia ?”

Emilie Jolivet, directrice de la division semi-conducteurs au sein de la société de conseil Yole Group, février 2024

- Concepteur *fabless* de puces électroniques, Nvidia est **spécialisé dans les processeurs graphiques** (GPU, pour *graphics processing unit*). Les équipements les plus puissants sont **généralement utilisés dans le domaine du jeu vidéo**, pour permettre l’affichage de graphismes de haute qualité. Ses technologies sont également présentes dans l’industrie du cinéma (réalisation des effets spéciaux). Entreprise américaine, Nvidia incarne le leader du segment.
- Fondée en 1993, la société s’est **imposée au fil du temps grâce à plusieurs innovations de rupture**, notamment l’invention du processeur graphique en 1999. Un an auparavant, elle s’est **associée au taïwanais TSMC** pour la fabrication de ses puces. Elle s’est également développée par le biais de **quelques acquisitions lui permettant d’obtenir des briques technologiques** : 3DFX en 2000, Media Q en 2003, Mellanox en 2020.
- Le groupe a toutefois diversifié les usages de ses GPU à partir de 2006, lorsque la découverte du calcul en parallèle a permis d’**employer les processeurs pour une grande variété de tâches**. Analyste chez IDC, Shane Rau expliquait début 2024 : “**Le parallélisme massif des GPU pouvait dès lors être exploité dans l’IA** et les mathématiques, et adapté pour des domaines spécifiques comme la modélisation météorologique ou la recherche de virus”. Nvidia a depuis lancé **de nombreuses innovations appliquées au domaine de**

l’IA, notamment en 2016 et en 2019. Supercalculateurs, Internet des objets, véhicules autonomes, robotique : les équipements de Nvidia se déploient pour une variété d’usages.

- Disposant de puissantes puces de 4 nanomètres produites par TSMC, l’entreprise bénéficie d’**une place incontournable** dans la fourniture d’équipements pour le développement de l’IA. L’entraînement de ChatGPT s’est ainsi appuyé sur 10 000 GPU H100 de Nvidia, rapporte *Challenges*. Meta compte obtenir de son côté jusqu’à 350 000 H100 et a installé le Research Supercluster, un supercalculateur composé de 16 000 GPU.
- Le déploiement à grande échelle des IA génératives a **propulsé les indicateurs financiers de Nvidia à des niveaux inédits**. En un an, son chiffre d’affaires a doublé et ses bénéfices ont été multipliés par sept. Le groupe affiche en outre une marge record de 50 %.
- Porté par ses performances commerciales et par les marchés financiers, **Nvidia a franchi plusieurs seuils symboliques**. Le groupe a pris place parmi les cinq premiers acteurs des puces, et a rejoint plusieurs GAFAM et Saudi Aramco (compagnie pétrolière saoudienne) parmi les valorisations les plus élevées de la planète, après avoir atteint les 1 000 milliards de dollars de capitalisation boursière en mai 2023.
- Ces résultats sont **le reflet d’un positionnement idéal** dans le nouveau contexte de marché, mais également **des sommes gigantesques que les groupes du numérique sont prêts à dépenser** pour s’imposer dans le domaine des IA génératives. Cette situation de dépendance pourrait cependant ne pas perdurer ; aussi, certains analystes se questionnent sur une éventuelle surévaluation du cours boursier actuel de Nvidia.

Des alternatives émergent pour réduire l'hégémonie de Nvidia

Les GAFAM développent leurs propres puces

“Si les hyperscalers [les grands acteurs du cloud] se fournissent eux-mêmes en composants, ils vont gagner des marges de négociation et pourront tirer les prix vers le bas.”

Emilie Jolivet, directrice de la division semi-conducteurs au sein de la société de conseil Yole Group, février 2024

- Percevant les risques d'**une trop grande dépendance à l'égard de Nvidia** et souhaitant réduire les coûts liés à leurs services d'IA, les grandes entreprises du numérique conçoivent **de nouvelles puces pour un usage interne** :

- **Amazon a lancé Trainium**, une puce spécifiquement dédiée à l'entraînement de modèles d'IA. Elle se veut plus rentable et performante par rapport aux solutions d'Amazon ba-

sées sur des GPU traditionnels. Trainium est intégrée dans les outils cloud d'Amazon et enrichit donc les prestations offertes par la division Web Services (AWS) ;

- **Microsoft travaille sur le projet Athena**, une initiative visant à créer une puce spécialisée dans l'IA pour s'affranchir de la dépendance à Nvidia. Ces modèles sont conçus pour **réduire les coûts et optimiser les temps de calcul** dans les data centers, avec l'intention de les rendre disponibles pour les projets internes ainsi que ceux en collaboration avec OpenAI ;

- Google a de son côté développé les puces TPU ;

- Meta a mis au point la puce MTIA.

- Malgré les revendications des GAFAM, les spécialistes indiquent que **les dernières puces de Nvidia demeurent bien plus puissantes**. Également présent dans les services de cloud pour l'IA, Nvidia représente donc pour les GAFAM un partenaire stratégique, un fournisseur coûteux et un concurrent.

Des start-up se positionnent avec de nouvelles puces

“Sur le plan technique, plus l’architecture de calcul est spécialisée, et plus elle est efficace mais plus ses domaines d’application sont restreints. Les GPU de Nvidia sont des processeurs génériques qui s’adaptent à n’importe quelle application. (...) La puce de Groq est spécialisée et ne fait que l’inférence, pas l’apprentissage.”

Marc Duranton, expert senior en systèmes embarqués et réseaux de neurones au CEA-List, mars 2024

- De jeunes sociétés cherchent à se positionner dans la conception et la commercialisation de puces spécialisées pour l’IA. En adoptant une telle approche, elles souhaitent proposer **des puces plus performantes et plus optimisées** car spécifiquement dédiées à l’IA. **Elles ciblent en général l’inférence**, soit l’exploitation des modèles de langage, l’apprentissage algorithmique nécessitant des puissances de calcul trop importantes.
- La société américaine **Groq**, fondée en 2016, a par exemple mis au point une puce de 14 nanomètres fabriquée par GlobalFoundries. Si ce degré de miniaturisation semble peu avancé comparé aux 4 nanomètres de Nvidia, la start-up compense cet écart par **une architecture différente et plus efficace**. Elle souhaite en outre parvenir à des tailles de puces plus petites à l’avenir.
- Faire produire par GlobalFoundries aux États-Unis peut constituer **un atout pour la société, en revendiquant une certaine souveraineté nationale** face à Nvidia, qui s’approvisionne à Taïwan.

- **Groq fortement développée depuis ses débuts**, avec un chiffre d’affaires de 60 millions d’euros en 2023. L’entreprise a lancé GroqCloud pour mettre à disposition ses puces LPU (*Language processing unit*) sous la forme d’un **service cloud**. La division Groq Systems a par ailleurs été ouverte pour **diversifier les marchés ciblés**.

- Grâce à **des levées de fonds d’un montant total de 367 millions de dollars** (dont 300 millions en série C en 2022), Groq a pu réaliser **deux acquisitions dans son secteur** : Maxeler Technologies en 2022, puis Definitive Intelligence en 2024.

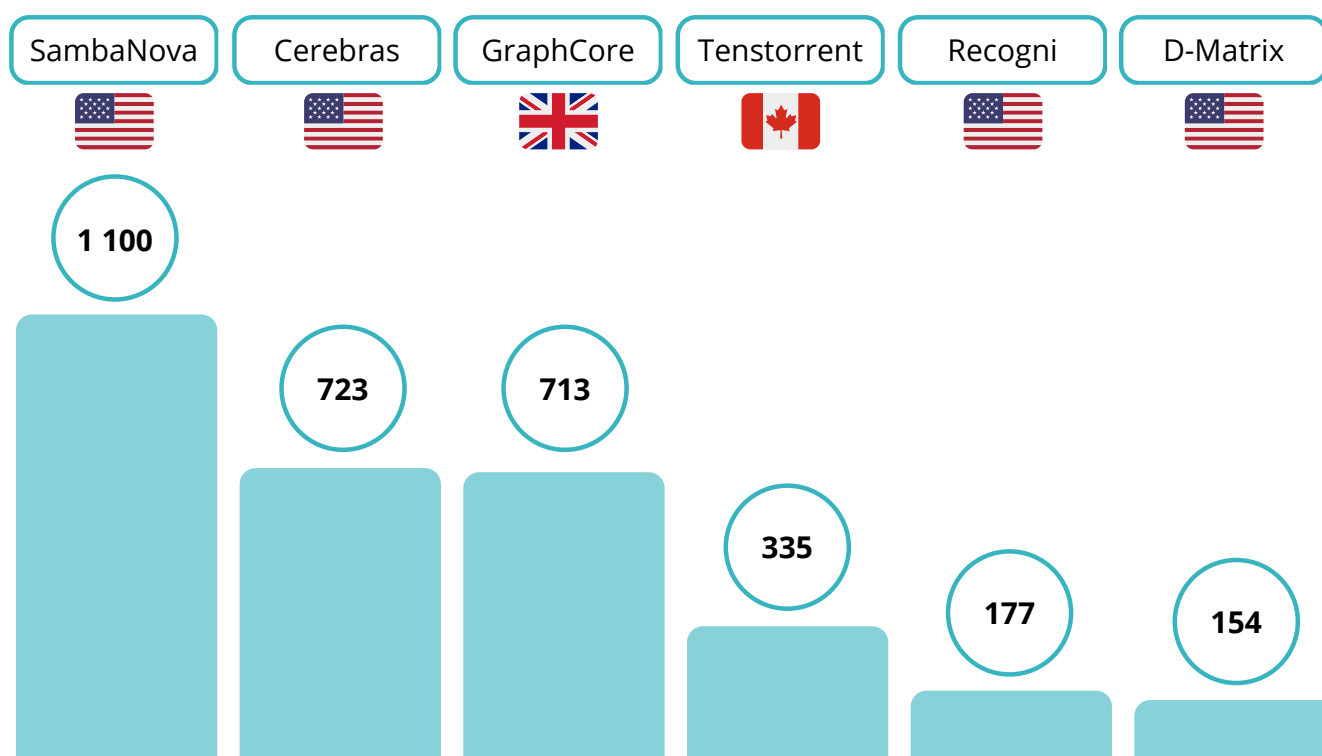
- Bousculer la domination de Nvidia dans les puces pour l’IA demeure toutefois complexe. Malgré les revendications de la start-up, **le rapport coût/performances ne lui est pas forcément favorable** : moins chères, ses puces doivent également être rassemblées en plus grand nombre pour égaler la puissance des GPU de Nvidia. **Ce dernier bénéficie d’économies d’échelle** dont Groq ne peut encore profiter, faute de volumes suffisants.

- **D’autres acteurs essayent de s’implanter** sur le marché, comme Cerebras, Graphcore ou encore SambaNova. En 2023, plus de 190 start-up dans le monde se positionnaient dans le développement de puces électroniques spécialisées dans les besoins des IA, selon la banque d’investissement Woodside Capital Partners.

- Plus simple à investir, **le segment de l’inférence, en particulier avec le modèle *fabless*** (“sans usine”, centré sur la conception et la commercialisation des puces), va de plus **constituer une part croissante du marché**, au fur et à mesure du déploiement des IA. Il représentera les deux tiers du secteur à l’horizon 2034 (source : IDTechEx).

Les financements affluent pour les start-up axées sur les puces pour IA

Fonds levés, en millions de dollars



Traitement IndexPresse.

Parmi les fonderies, TSMC apparaît comme le leader incontesté des puces premium

Des parts de marché et des investissements largement supérieurs à la concurrence pour TSMC

Parts de marché par type de puces, en 2020

- 130 nanomètres et plus 30 %
- entre 32 et 12 nanomètres 70 %
- entre 10 et 5 nanomètres 90 %

au quatrième trimestre 2020

part du chiffre d'affaires

- 7 nanomètres 29 %
- 5 nanomètres 20 %

Investissements

- 100 milliards de dollars sur la période 2021-2023
- 28 milliards de dollars supplémentaires en 2021
- 32 milliards de dollars minimum prévus en 2024

accroître les capacités de production
+
poursuivre la recherche

Traitement IndexPresse. Sources : Bain, IC Insights, Gartner, *Le Comptoir du hardware*

“Le parc de Hsinchu rassemble les meilleures entreprises de conception et de fonderie de puces électroniques de Taïwan, ainsi que les meilleures universités et instituts de recherche. Cet effet cluster est au cœur de notre modèle.”

Chen Shu-Chu, directrice du parc de Hsinchu, février 2022

- L'explosion des besoins en IA bénéficie fortement aux fabricants de puces, qu'ils s'adressent à Nvidia, aux GAFAM ou à des start-up du domaine.

- **Le taïwanais TSMC** (Taiwan Semiconductor Manufacturing Company) constitue **le principal fondeur à l'échelle mondiale**, avec une part de marché variant entre 55 % et 60 % selon les années (sources : *Challenges*, Statista). Samsung, en seconde place, reste loin derrière avec environ 18 % de parts.

- **Son positionnement affirmé dans les semi-conducteurs les plus performants** avantage TSMC dans le contexte actuel. Hégémonique

dans la production des plus petites puces, il en retire en outre des bénéfices substantiels (38 % de marge nette en 2023). Entreprise cotée la mieux valorisée d'Asie, elle a connu une hausse de 32 % de son cours de Bourse durant l'année 2023.

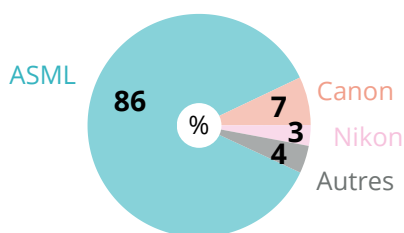
- **La société bénéficie d'économies d'échelle élevées** grâce à sa production importante (100 000 *wafers* par mois à Taïwan, contre 20 000 pour son projet d'usine en Arizona, aux États-Unis.) Elle est capable de **produire des puces de 3 nanomètres, tandis que celles de 2 nanomètres sont en développement**. Elle a toujours été pionnière en termes de miniaturisation, quelle que soit la catégorie (première dans les puces de 7 nanomètres en 2018, dans celles de 5 nanomètres en 2020).

- Son avance dans la recherche résulte **d'investissements massifs et constants**. Par conséquent, les autres fonderies se sont elles aussi lancées dans des plans d'investissements majeurs. Samsung a ainsi prévu d'investir 116 milliards de dollars sur la décennie 2021-2030. Des évolutions pourraient toutefois voir le jour si certains marchés (automobile, produits électroniques, smartphones) venaient à ralentir.

Les progrès technologiques se poursuivent dans les équipements de lithographie

Les équipements de lithographie, un marché en monopole

Parts de marché mondiales, en 2022



Traitement IndexPresse. Source : Yole Développement

Marché mondial des équipements de lithographie

17,8 milliards de dollars en 2022

Marché mondial des équipements de lithographie par nano-impression

57 millions de dollars en 2022

“La lithographie par nano-impression va probablement rester pendant encore des années au stade du laboratoire ou de la production pilote. (...) Je ne pense pas que cette technologie brise le monopole du néerlandais d’ici cinq ans.”

Taguhi Yeghoyan, analyste au cabinet Yole Développement, janvier 2024

- En lien avec les fonderies, les équipementiers investissent pour miniaturiser encore davantage les puces électroniques et permettre d’accroître les gains de performance. TSMC et Samsung comptent atteindre 2 nanomètres en 2025 et 1,4 nanomètre en 2027. De son côté, Intel vise 1,8 nanomètre.

- Le marché des équipements de lithographie, qui permettent la gravure des semi-conducteurs, se montre fortement concentré : l’entreprise néerlandaise ASML en capte la quasi-totalité.

- ASML a pour ambition d’atteindre 1 nanomètre, voire de proposer des équipements dédiés aux puces subnanométriques, en utilisant une lithographie à UV extrêmes à haute couverture numérique. Développée en laboratoire, la technologie se révèle pour l’instant très onéreuse : une seule machine de ce type coûte 350 millions d’euros.

- ASML a approfondi son partenariat avec le centre belge Imec pour développer une ligne pilote de lithographie avancée pour les puces nanométriques.

- En France, le CEA-Leti investit 500 millions d’euros, majoritairement financés par le plan France 2030, pour miniaturiser les puces FD-SOI et passer de 22 nanomètres à 10 nanomètres. Elles sont notamment utilisées dans l’Internet des objets et les produits électroniques.

- Le japonais Canon, challenger d’ASML dans la lithographie, mise sur la technologie de la nano-impression. Par rapport à la norme de l’UV, elle présente divers avantages, notamment un prix plus bas. Une machine de Canon coûte ainsi cinq à six fois moins cher qu’un équipement d’ASML.

L’objectif est de concurrencer ASML sur les puces de 7 nanomètres et moins, où ce dernier est aujourd’hui en situation de monopole. Canon viserait d’abord 5 nanomètres avant d’étendre sa technologie aux puces de 2 nanomètres. Les risques accrus de défauts peuvent toutefois pénaliser cette technologie pour les grands volumes.

Entre opportunités et menaces, le contexte géopolitique guide les mutations du secteur

Les gouvernements à la manœuvre pour promouvoir la production locale

“Avant la pénurie, notre industrie avait peu de contacts avec la politique. On ne faisait que du business. Mais cela a changé ces derniers mois. Tous les pays viennent nous voir pour nous dire qu’ils veulent leurs propres usines de semi-conducteurs !”

Terry Tsao, président de l’association Semi Taïwan, février 2022

“Tant que nous dominerons la production de semi-conducteurs, personne n’aura intérêt à ce que la Chine nous envahisse.”

Frank Huang, président de l’entreprise Powerchip, février 2022

- Le développement de TSMC a été, dès l’origine, **le fruit d’une stratégie gouvernementale taïwanaise**. L’entreprise a été créée par l’Institut de recherche en technologie industrielle, un laboratoire public. Ce dernier a **permis de fonder plus de 240 sociétés**. La mise en place de parcs industriels comme celui de Hsinchu a contribué à **la construction d’un écosystème**, qui compte maintenant dans ses rangs un acteur comme ASE, leader mondial du test et de l’assemblage de puces électroniques. L’industrie des semi-conducteurs constitue à présent **un secteur stratégique** pour le pays.

- Les autres gouvernements tentent de **soutenir leur filière nationale** et d’accompagner les plans d’investissement mis en place par les entreprises. Ces politiques ont été **accélérées par la pénurie** à la suite de la crise sanitaire de 2020 : les perturbations de la chaîne d’approvisionnement ont **révélé les fortes dépendances** de certains pays aux importations.

- Dans la commercialisation, le marché connaît en outre **une mutation depuis une vingtaine d’années**, avec la montée en puissance de la Chine, et dans une moindre mesure de la Corée du Sud, au détriment des États-Unis et du Japon. Chacun cherche à rester compétitif dans cet environnement concurrentiel de plus en plus rude.

- Des plans de relocalisation ont été lancés, à commencer par **le Chips Act américain**. Doté de 52 milliards de dollars, il aurait généré 300 milliards d’investissements privés, avec un objectif de 500 milliards de dollars à terme. Parmi les multiples sites prévus, **l’installation d’une usine de TSMC en Arizona** concerne la fabrication de puces de 5 nanomètres, pouvant être utilisées pour les applications d’IA. **Doté de 40 milliards de dollars, il s’agit du plus gros investissement étranger** de l’histoire des États-Unis.

- TSMC devrait également ouvrir **une usine dédiée aux puces pour IA à Kumamoto**, au Japon. L’investissement s’élève à 8,6 milliards de dollars.

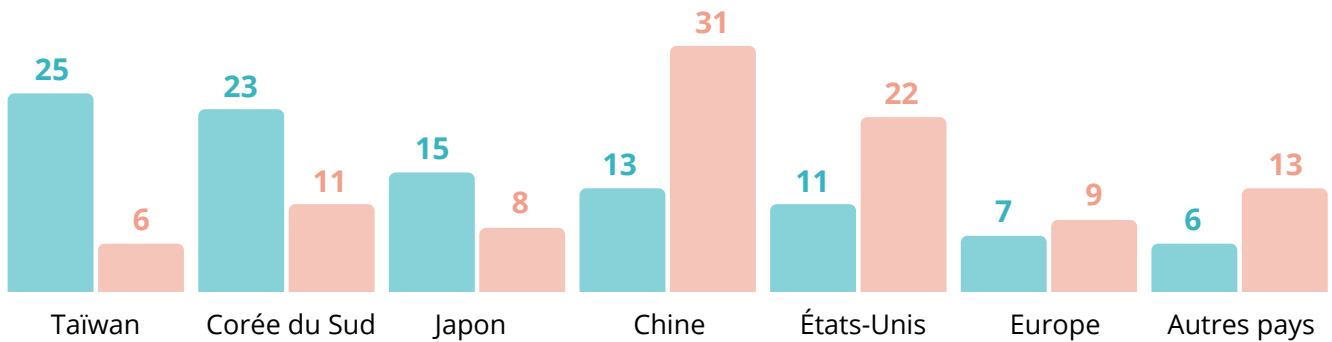
LA FILIÈRE DES PUCES ÉLECTRONIQUES PROFITE DE CE NOUVEL ÉLAN

- La Corée du Sud a lancé un **K-Chips Act**, un plan de 450 milliards de dollars sur vingt ans, s'ajoutant aux 230 milliards prévus par Samsung. Le gouvernement a également mis en place des **exonérations fiscales** : 15 % sur les installations industrielles des *chaebols* (conglomérats) et 10 % sur les investissements. Ces aides visent à **contrebalancer le crédit d'impôt de 25 %** qu'accorde sans conditions Taïwan à ses grandes entreprises.

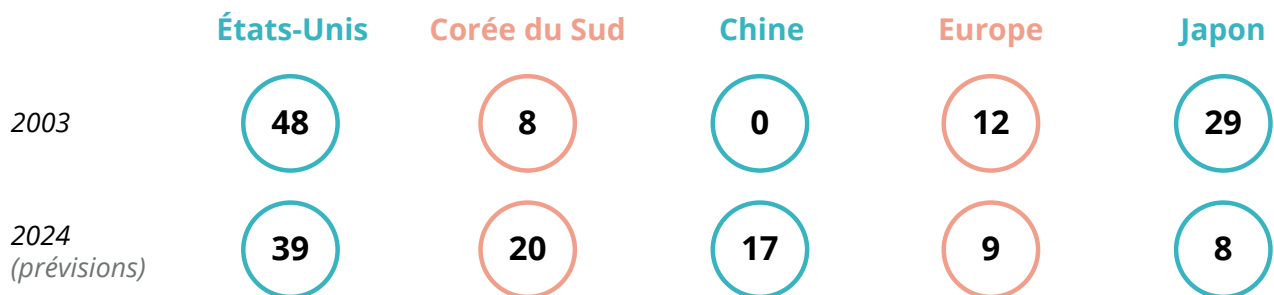
- L'Europe dispose elle aussi de son propre programme, l'**European Chips Act**. Il affiche un budget de 43 milliards d'euros et vise à **atteindre une production représentant 20 % du marché mondial** en 2030, contre 9 % en 2022.

Des dépendances aux importations et une concurrence qui s'affirme dans la commercialisation de puces

Production et consommation mondiales de puces, tous types confondus, par pays, en % (données disponibles en mai 2023)



Parts de marché mondiales dans les ventes de puces, en %



Traitement IndexPresse. Sources : L'Usine nouvelle, Challenges

Des restrictions antichinoises aux conséquences incertaines

“Les États-Unis disent vouloir freiner le développement militaire de la Chine. Mais avec leurs dernières sanctions, ils vont bien au-delà, avec le risque d’un effet collatéral important sur les applications civiles.”

*Mathieu Duchâtel, directeur du programme
Asie à l’Institut Montaigne, mai 2023*

“[Les États-Unis] mènent une intense diplomatie pour convaincre leurs alliés de restreindre l’exportation de technologies de semi-conducteurs vers la Chine.”

*Mathilde Velliet, chercheuse en géopolitique
des technologies à l’Institut français
des relations internationales (Ifri), mai 2023*

- Soulignant les usages duales (civiles et militaires) des semi-conducteurs, **les États-Unis ont mis en place des mesures visant à pénaliser l’industrie chinoise** dans le domaine et à limiter les possibilités d’importations du pays en semi-conducteurs. Fin 2022, ils ont ainsi **interdit l’exportation vers la Chine des puces les plus sophistiquées**, utilisées dans l’IA et dans les technologies militaires. Auparavant ciblées sur des entreprises, les mesures visent à présent la Chine dans son ensemble ainsi que les résidents américains. **Par son caractère extraterritorial, elle concerne aussi les fournisseurs étrangers** dont les produits incorporent des composants américains.

- Les États-Unis ont également poussé les Pays-Bas et le Japon, qui dominent le marché des machines de lithographie extrême ultraviolet, à **bannir l’exportation de ces équipements** spécialisés vers la Chine. Le néerlandais ASML est notamment le seul fournisseur de machines pour les puces de 7 nanomètres et moins.

- Le pays s’est en outre rapproché de Taïwan, de la Corée du Sud et du Japon pour **bâtir l’alliance Chip 4**. En maîtrisant à eux quatre la totalité de la chaîne de valeur, l’objectif est de développer de nouvelles technologies de puces.

- Ces différentes mesures font réagir la Chine : celle-ci aurait **dérobé des secrets technologiques** en février 2023 grâce à un employé d’ASML. Comme le souligne un fournisseur de TSMC, ce type de pratiques est un risque permanent dans le domaine : “C’est un secteur très secret, où chacun cherche à voler les innovations des autres”.

- Les initiatives américaines peuvent **engendrer des effets ambigus**. La fermeture du marché chinois pénalise de nombreux acteurs, y compris aux États-Unis. La fédération américaine du secteur a ainsi clairement critiqué les annonces gouvernementales. Celles-ci pourraient en outre **se montrer contre-productives**, forçant la Chine à investir dans toute la filière et à améliorer sa maîtrise technologique, tant au niveau des équipements que de la conception. **L’alliance Chip 4 suscite en outre une méfiance** chez ses membres, qui ne veulent pas perdre leur avance au profit des États-Unis.

- Certains acteurs du marché peuvent toutefois **profiter de cette rivalité sino-américaine**. L’utilisation de semi-conducteurs chinois étant proscrite, les entreprises américaines doivent **se fournir auprès des fabricants sud-coréens ou taïwanais**. La concurrence s’intensifie entre eux pour capter le marché. L’émergence de sites industriels dans une logique de relocalisation **bénéficie aux équipementiers**. Aux États-Unis, la fabrication sur le sol national peut représenter **un argument commercial pour les concepteurs de puces** : c’est l’approche suivie par la start-up Groq pour se démarquer de Nvidia et de son approvisionnement taïwanais.

Une Europe globalement absente dans la fabrication des puces les plus avancées

Des projets centrés sur des puces moins performantes

- Malgré les fonds alloués au European Chips Act, **l'Europe semble en retard dans le secteur**. Si elle possède, avec ASML, une place incontournable dans les équipements, elle ne dispose pas de sites de fabrication de puces ultra-sophistiquées. Ses usines peuvent produire au mieux **des puces entre 22 et 18 nanomètres**, comme celles fabriquées par STMicroelectronics à Crolles (Isère).

- Cette situation se prolonge dans le cadre des nouveaux projets, **notamment en Allemagne où se multiplient les implantations**. Intel installe une usine à Magdebourg, financée à hauteur de plus de 10 milliards d'euros par l'Allemagne ; TSMC porte un projet à Dresde avec Bosch, NXP et Infineon. Ce dernier y construit également une nouvelle usine. Concentrant les capitaux publics et privés, ces sites de fabrication portent sur **des puces de taille plus importante**, parfois en micro-mètres.

- Les marchés ciblés par ces usines s'avèrent donc différents de ceux visés par les puces les plus abouties. **Elles s'adressent surtout à l'automobile** et à divers produits industriels, mais leurs puces ne sont pas adaptées pour le calcul à haute intensité utilisé dans l'IA.

- Plusieurs raisons expliquent ce phénomène. D'une part, **les pénuries ont été marquantes pour le secteur automobile**, qui a souffert de ces dysfonctionnements logistiques et affiche un poids considérable sur le continent. Engagés dans le développement de l'électrique, une motorisation nécessitant davantage de semi-conducteurs, **les constructeurs préfèrent développer un approvisionnement européen**. Ils sont ainsi à la manœuvre – ou y ont été poussés par leurs États – dans la plupart des projets d'usines de batteries électriques, futures clientes des fabricants de puces.

- De leur côté, les fonderies cherchent à **se démarquer des acteurs tels que TSMC**, dont l'avance sur les puces les plus performantes paraît irrattrapable. **La concurrence est plus ouverte sur un marché comme l'automobile**, à la fois en croissance et nécessitant des puces moins sophistiquées.

- Si l'essor de l'IA générative a stimulé la demande en puces électroniques chez les GAFAM, **le manque de clients de cette envergure en Europe freine également le développement** de sites industriels dédiés. Les gouvernements focalisent donc les subventions sur l'approvisionnement de filières d'importance déjà existantes.

Des start-up françaises s'engagent dans les puces pour IA

“Chaque révolution technologique rebat les cartes du marché : Intel avait dominé le marché des processeurs de PC, Qualcomm celui des processeurs de smartphones, Nvidia celui du cloud. Nous voulons être le leader de la future rupture, l'intelligence embarquée, ou edge computing.”

Éric Baissus, président du directoire de Kalray, décembre 2020

- Quelques sociétés françaises se positionnent néanmoins dans les puces pour l'IA. Fondée en 2008, Kalray a développé **des processeurs pour les data centers et pour l'edge computing**, domaine où elle vise le leadership mondial. Il s'agit d'**un traitement de données décentralisé**, au plus proche du terrain, adapté aux systèmes d'IA tels que la robotique, les voitures autonomes, l'armement, etc. **Kalray est un fabless** (concepteur sans usine) et fait produire ses puces (16 nanomètres) à Taïwan et aux États-Unis. L'assemblage des puces pourrait toutefois être rapatrié en France.

- GreenWaves Technologies a quant à elle mis au point **des puces optimisées pour l'IA embarquée** : sa faible consommation énergétique lui

permet de servir dans des objets connectés fonctionnant sur batteries, notamment les caméras intelligentes. Elle a **levé 32 millions d'euros** depuis sa création, dont 20 millions en 2023.

- Grai Matter Labs, fondée en 2016, se focalise sur **la conception de puces optimisées dans une logique de calcul neuromorphique**. Rapprochant calcul et mémoire, elle peut accélérer le traitement des données tout en améliorant l'efficacité énergétique. Sa dernière puce, GrAI VIP, de 12 nanomètres, est fabriquée aux Pays-Bas. La société a **levé 13 millions d'euros** en 2020 et a enregistré 1 million d'euros de précommandes en 2022.

- Une autre société, Hawai.tech, s'est lancée en 2019 sur **le segment des puces d'IA misant sur la frugalité et l'edge computing**. En 2023, elle menait des expérimentations dans le domaine de la maintenance prédictive.

- Née en 2020, la start-up Wise-Integration travaille avec TSMC sur les matériaux composites. Issue du CEA-Leti, elle a créé **un matériau pouvant transporter 1 000 fois plus d'électrons que le silicium**. Son innovation permet d'améliorer la performance et la miniaturisation des alimentations, notamment des data centers.

PRINCIPALES SOURCES UTILISÉES

- Beky Ariane, "Intelligence artificielle : les investissements s'accélèrent", *silicon.fr*, 9 août 2021
- Belhassen Antoine, "Nouvelle usine STMicroelectronics : un projet de 7,5 milliards d'euros, un des plus grands investissements industriels des dernières décennies", *francetvinfo.fr*, 5 juin 2023
- Bergounhoux Julien, "Meta regroupe ses divisions de recherche en IA et se vante d'avoir l'équivalent de 600 000 Nvidia H100", *usine-digitale.fr*, 19 janvier 2024
- Bouchaud Bastien, "Nvidia franchit le seuil des 2 000 milliards de dollars en Bourse", *lesechos.fr*, 3 mars 2024
- Chardenon Aude, "Amazon lance Trainium, une nouvelle puce conçue pour entraîner des modèles de machine learning", *usine-digitale.fr*, 3 décembre 2020
- Cheminat Jacques, "Semi-conducteurs : l'Europe mobilise 22 Md€ pour plusieurs projets", *lemondeinformatique.fr*, 9 juin 2023
- Coutures Alix, "Nvidia dicte sa loi au cœur de la machine IA", *Challenges*, 15 février 2024, p.54-57
- Filippone Dominique, "Le marché européen de l'IA plus fort que la crise", *lemondeinformatique.fr*, 21 février 2023
- Floretti Célio, "La Corée du Sud achète sa place dans la course aux puces", *Challenges*, 16 février 2023, p.44-45
- Grésillon Gabriel, "Google investit 2 milliards de dollars dans Anthropic, concurrent de ChatGPT", *lesechos.fr*, 29 octobre 2023
- Le Denn Arthur, "Vers un retour en grâce des semiconducteurs après plusieurs revers", *Investir - le Journal des Finances*, 10 décembre 2022, p.14-15
- Leblal Serge, "STM et GlobalFoundries s'allient pour produire des puces à Crolles", *lemondeinformatique.fr*, 11 juillet 2022
- Loukil Ridha, "La guerre des puces gagne la planète", *L'Usine nouvelle*, mai 2023, p.118-120
- Loukil Ridha, "Canon s'attaque au monopole d'ASML", *L'Usine nouvelle*, janvier 2024, p.94-96
- Loukil Ridha, "Comment STMicro tente de maîtriser sa soif", *L'Usine nouvelle*, avril 2023, p.108-109
- Loukil Ridha, "Kalray, la pépite française qui se rêve en Nvidia européen, prend son envol", *usinenouvelle.com*, 24 janvier 2022
- Loukil Ridha, "La start-up grenobloise GreenWaves lève 20 millions d'euros pour produire sa puce d'IA embarquée", *usinenouvelle.com*, 2 février 2023
- Loukil Ridha, "Qui est Groq, cette start-up californienne qui veut défier Nvidia dans les puces d'IA ?", *usinenouvelle.com*, 6 mars 2024
- Loukil Ridha, "Cap sur les puces subnanométriques", *L'Usine nouvelle*, septembre 2023, p.14-15
- Mabille Philippe, "Nvidia Wars : celui qui contrôle les puces contrôle l'univers !", *latribune.fr*, 3 mars 2024
- Mens Yann, "Les États-Unis veulent garder leurs puces", *Alternatives Économiques*, décembre 2022, p.50
- Montflier Frédéric, "Électronique neuromorphique. Contenir l'envolée énergétique de l'IA", *Industrie & Technologies*, septembre 2023, p.28-43
- Muthoni Edith, "Samsung Semiconductor Revenue Market Share Dips by 40% in 2 Years", *stocklytics.com*, 14 février 2024
- Poitiers Mélicia, "Projet Athena : Microsoft prépare une puce spécialisée dans l'IA pour s'émanciper de Nvidia", *usine-digitale.fr*, 19 avril 2023
- Séramon Célia, "IA générative : Groq acquiert Definitive Intelligence pour faire passer sa plateforme cloud à la vitesse supérieure", *usine-digitale.fr*, 4 mars 2024
- Simorre Adrien, "Taiwan, géant convoité des puces électroniques", *L'Usine nouvelle*, février 2022, p.49-55
- "Semi-conducteurs : le Taïwanais TSMC accuse une lourde chute de son bénéfice", *latribune.fr*, 19 octobre 2023

La collection IndexPresse *Business Etude*

Comment accéder à des données fiables, pertinentes et surtout synthétisées, alors que l'information n'a jamais été aussi accessible en apparence ?

Voilà une question à laquelle sont confrontés quotidiennement les décideurs dans les entreprises lorsqu'il s'agit de prendre les bonnes décisions.

C'est pourquoi nous avons créé la collection **IndexPresse Business Etude**, des études sectorielles complètes, réalisées à partir des plus grands titres de la presse

économique et professionnelle. En s'appuyant sur des informations fiables et de qualité, les études d'IndexPresse offrent des synthèses analytiques et éclairées sur les secteurs d'activité émergents ou en mutation.

Vous aurez ainsi toutes les clés en main pour accompagner votre réflexion stratégique, en vous appuyant sur l'examen des enjeux de votre marché, afin d'anticiper ses évolutions et valider, ou modifier, votre positionnement dans le jeu concurrentiel.

IndexPresse *Business Etude*

Date de parution - mars 2024.



Renaud HAMMAMY

renaud.hamamy@indexpresse.fr

Auteur

SECTEURS & MARCHÉS
PUCES ÉLECTRONIQUES ET IA
MARS 2024

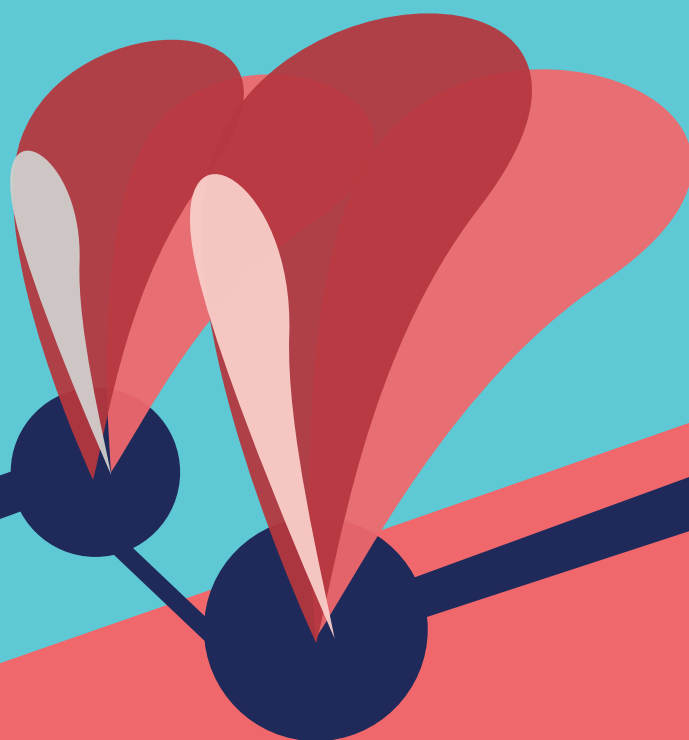


Photo de couverture : © Graphic Leading



IndexPress

IndexPress
19 rue René Thomas
38000 Grenoble
Tél. 04 76 92 05 25
indexpresse@indexpresse.fr